Fermi National Accelerator Laboratory

FERMILAB-Conf-92/282

# Drift Chamber Tracking with a VLSI Neural Network

Clark S. Lindsey, Bruce Denby and Herman Haggerty

*Fermi National Accelerator Laboratory*
*P.O. Box 500, Batavia, Illinois 60510*

Ken Johns

*Dept. of Physics, University of Arizona*
*Tucson, Az. 85721*

October 1992

Invited talk at *Second Workshop on Neural Networks: From Biology to High Energy Physics*, Marciana Mariana, Isola d'Elba, Italy, June 18-26, 1992

Operated by Universities Research Association Inc. under Contract No. DE-AC02-76CHO3000 with the United States Department of Energy

## Disclaimer

*This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.*

# Drift Chamber Tracking with a VLSI Neural Network[*]

**Clark S. Lindsey**[†] **Bruce Denby, Herman Haggerty**
*Fermi National Laboratory*[‡]
*P.O. Box 500, Batavia, Il. 60510, USA*

**and**

**Ken Johns**
*Dept. of Physics, University of Arizona*
*Tucson, Az. 85721, USA*

We have tested a commercial analog VLSI neural network chip for finding in real time the intercept and slope of charged particles traversing a drift chamber. Voltages proportional to the drift times were input to the Intel ETANN chip and the outputs were recorded and later compared off line to conventional track fits. We will discuss the chamber and test setup, the chip specifications, and results of recent tests. We'll briefly discuss possible applications in high energy physics detector triggers.

## 1. Introduction

Neural networks implemented in VLSI offer the capability of very fast pattern recognition. The first generation of such chips can process an entire multilayer network in microseconds and could provide on line processing. In high energy physics experiments, an important on line application of VLSI neural networks would be the selection of events written to tape. (Here an event refers to a collision of two particles and the subsequent signals induced in the detectors from the reaction products.) This on line selection process, called triggering, is a crucial part of an experiment. Collisions of particles can occur at rates of hundreds of kHz and the rate will include background interactions such as stray beam particles striking beam pipe material. Usually the most interesting events occur only at a small fraction of the total rate and also the recording the detector signals onto tape for later off line processing is typically limited to tens of Hz. So the trigger system must select only those types of events of greatest interest, while efficiently rejecting the backgrounds.

The first step or level in a typical trigger system makes cuts on simple raw signals such as the amplitude of the signal from a calorimeter. The second level of the trigger transforms the raw signals from the detectors into more elaborate quantities such as the trajectory parameters

[*] Invited talk given at *Second Workshop on Neural Networks: From Biology to High Energy Physics*, Marciana Mariana, Isola d'Elba, Italy, June 18-26, 1992

[†] Current address: 1261 Foran Ln., Aurora, Il. 60506

[‡] Fermilab is operated by the Universities Research Association under contract with the Department of Energy.

of the particles and then makes cuts on these quantities. Each level in the trigger process will reduce the number of accepted events passed to the next level by as much as factors of 10-100. Typically, the first two trigger levels are in hardware. After the second level the trigger rate has been reduced to a rate that can usually be handled by software. After perhaps two more trigger levels the remaining events are written to tape.

The use of hardware neural networks in high energy physics has been discussed frequently and several projects are currently in development.[1,2] Here we discuss using a VLSI neural network chip to find the slope and intercept of a charged particle traversing one type of commonly used particle detector called the drift chamber. This technique could be useful for a second level trigger where, for example, there might be a requirement that a minimum number of tracks point to the same collision vertex. The drift chamber provides analog voltage signals proportional to coordinates along the trajectory of the particle through the chamber's active volume. The chip receives these signals as inputs and provides the slope and intercept of the track as output. We reported previously on our initial results[3]. Here we review those results and report on some subsequent developments.

## 2. Drift Chambers and Test Setup

Figure 1a shows a cross-section of a single drift chamber cell. The cell consists of a sense wire (typically $30\mu m$ in diameter) at a high positive voltage. The wire is inside of an aluminum channel held at ground potential. The channel is filled with a gas such as a CO2/Ar mixture. If a charged particle traverses the gas, it will ionize some of the gas molecules along its path. The electrons will *drift* towards the positive voltage sense wire and the ions will drift towards the grounded aluminum. When the electrons reach the high electric field close to the wires there is an avalanche of ionization that produces a current pulse detectable by an amplifier attached to the wire end. The time between when the particle initially crossed the cell (this start time can be known from the accelerator timing or from another detector) and the time of the subsequent wire pulse is proportional to the distance between the wire and the point of closest approach of the track. The drift velocity of the electrons for the chamber used here is about $5cm/\mu s$. The sense wire is 5cm from the right and left walls and so the maximum drift time is about $1\mu s$. A time to voltage converter (TVC) was used here to produce a voltage proportional to the drift distance.

Note that there exists a left-right ambiguity. One doesn't know from a single cell signal which side of the wire the particle passed. Figure 1b shows a section of a 4-layer drift chamber. The cells are staggered with respect to each other so that the pattern of hits can eliminate the ambiguities.

For the particular chamber used here, the cells are actually paired together in that the sense wires are electrically connected at one end (see figure 1c). This allows one to measure the drift time and also the transit time for the signal to travel along the wire (difference in arrival time of pulses at each end). The transit times are of the order of a few nsecs for the meter long chamber used here. No correction to the drift time measurements was made for tracks in the left versus right cell or for the variation in transit times due to variation in track position along the wire.

As seen in figure 1a, above and below the wires there are metal pads. The ionization avalanche at the wire induces signals onto these pads for use, in combination with the signal transit times, to do a finer measurement of the track position along the wire. For our use, these analog pad signals were used to set a latch of 3v output if the signal was above a threshold. Here we only use the pad latch signal of the right hand cell of each pair to determine which of the two cells was hit. For example, if the pad signal of the right hand cell was 3v, then the track was in that cell, but if it was 0v, the track was in the left hand cell.

The chamber design illustrated in figure 1 is that of the chambers used to detect muons in the D0 experiment at the proton-antiproton collider at Fermilab.[4,5] These chambers can have up to 96 cells and be several meters long. For these test, however, a small prototype was used that was about a meter long and had only eight cells in four layers (three layers were used in ref. 3). Figure 2 shows the test setup. For the data here, cosmic rays were used in place of beam particles. Coincidences of two scintillator counters act as the trigger. The TVC and pad latch signals are sent to the neural network chip, which is then read out from an analog bus. A computer reads out the digitized values of the TVC, pad latches and neural net chip outputs. This chamber has a position resolution of about $500\mu m$. We define resolutions here as the sigma of Gaussian fits to distributions of residuals for fit tracks. For the neural net cases we use sigmas of Gaussian fits to distributions of differences between the parameters from the fit and the neural network. Since the neural network resolution is usually much larger than the fit resolution we ignore the error in the fit parameters.

## 3. Neural Network Architectures

We want to use the analog drift time voltages and the pad latches as inputs to a neural network which then computes the slope and intercept of the track. Here the slope is given as the angle from vertical (±45°) and the intercept is the crossing point in the plane of the sense wires of the second layer from the top (in figure 1b, 0 cm to 20 cm from the left cell wall of the left cell to the right hand cell wall of the right hand cell). The simplest method to implement the net consists of one output neuron having an activation proportional to the intercept, and a second neuron whose activation is proportional to the slope. However, there are several drawbacks to this scheme. As described in the next section, the chip used here has limited weight resolution (6-7 bits), a maximum weight of ±2.5, nonlinearities in the weight-input multiplications, and electronic jitter. These imperfections limit how accurately the chip can calculate the activation of a single neuron.

For some applications this type of proportional output scheme would still be desired but here we used an alternative method. Figure 2 shows the architecture of the network used. It is a feedforward network with 12 inputs, a hidden layer with 48 neurons, and an output layer with 64 neurons. The four drift time voltages are repeated twice (to help overcome the limited maximum weight) and there are four pad voltage inputs. The outputs are divided into two sections of 32 neurons each. The first section provides the intercept and the second section provides the slope. The intercept range of 0cm to 20cm is divided into 32 *bins* of 0.625cm. The first neuron represents the range 0cm to 0.625cm, the second represents 0.625cm to 1.25cm, etc. For a given intercept value the target distribution is a Gaussian (with sigma = 0.5cm) centered at the intercept point. This results in 3-4 neurons activated. There is a similar scheme for the slope section where each neuron represents a bin of 2.9° and ranges from ±45°. To determine the intercept, the neuron with the maximum activation is found and an average over ±2 neurons around it is calculated.

One advantage of this method of averaging over several neurons is a reduced sensitivity to jitter in a single neuron. Another advantage is that for cases where there are more than one possible track due to the left-right ambiguities, the output can show more than one answer (i.e. multiple *bumps*). The disadvantages of this *distributed* output method are the large number of neurons needed and the need for a second circuit to do the averaging.

Figure 4 shows a comparison of the proportional output method versus the distributed method using a simulation program. Here the maximum weight was limited to ±2.5 but there were no nonlinearities, electronic jitter, or limited precision included. Simulated track events were produced for the 4-layer chamber by sending tracks at random angles across the cells and requiring that the tracks cross active regions of one cell in each layer. A drift resolution of $500\mu m$ was simulated. Sets of 10000 training patterns were made, with each pattern having the network input and target output values. The neural network simulations each used 12 inputs values (2x4 drift times + 4 pad voltages). The simulations were done for 16, 32, 48, and 64 hidden units for both cases. The networks were trained with a back-propagation program (modified to limit the maximum weight) for greater than ten million iterations. The results in figure 4 were made with independent patterns not used in the training. One sees that the distributed output scheme gives about 2-3 times better resolution than the proportional output scheme. Implementation of some of these nets in the chip is currently being studied but the resolutions are expected to be roughly twice as big.

## 4. Neural Network Chip

The Intel Electrically Trainable Analog Neural Network (ETANN) chip is described in detail in references 6-8. Experience in using it for tracking was discussed in reference 3. We will only briefly review it here. The chip accepts up to 64 inputs and can be configured for 64 hidden neurons and 64 output neurons. Each neuron is connected to 16 internal fixed voltages (biases) so there are 80x64=5120 synapses for the hidden and the same for the output layer for a total of 10240 synapses. The synapse is basically a Gilbert multiplier circuit with two floating gates whose voltages determine the output current of the synapse for a given input voltage. Based on eprom technology, the charge on the gates can be altered by using large voltage pulses to cause electrons to tunnel through the insulation layer to the metal gate, where the charges will remain indefinitely. The gates are nonvolatile. The currents of 80 synapses are summed and presented to a threshold amplifier that represents the neuron. The amplifier turns on with a response similar to the sigmoidal characteristics desired.

A PC based system is available that lets one communicate with the chip. One can do emulation of the chip, loading and reading of weights, presentation of input voltage patterns and readout of the chip. The chip does not have learning circuitry built into it. Instead, for back-propagation, the PC calculates the delta weight changes for each iteration. This chip-in-the-loop (CIL) processing is fairly slow but doing the emulation first usually means that the CIL starts off fairly close to the desired performance.

## 5. Results

In reference 3 we discussed the use of the chip for finding tracks in three layer events. Figures 5a-b show typical tracks through the 3 layer chamber. The chip intercept resolution was $850\mu m$ and the slope resolution was $15mrad$. With only 3 layers there were often ambiguous tracks. Figure 5c-d show typical cases where the net chose a different track that the one obtained from the best fit. Notice, however, in the output distributions that there are bumps at the fit slope or intercepts, they just were not the largest ones. If one combined the the output of one chamber with another one, a secondary bump might become the largest one in the combined distribution.

Recently we have used the chip for a 4-layer chamber. Figures 6a-c show tracks in the four layer chamber. In the 3-layer chamber study there were trigger timing problems that prevented coverage over the full distance from the sense wire to the cell walls. We have improved the timing here so that there is full coverage. With the 4-layers there are reduced possibilities of ambiguous tracks but they still occur. Figures 6d shows one such case. However, the percentage of these is much less than for the 3-layer case (less than half percent in simulations.) Analysis of the four layer tracking is currently underway and resolutions and efficiencies will be reported later.

## 6. Discussion

We have shown that the Intel analog neural network chip can find track parameters from drift chamber signals in real time. This work was done in a test setup. We hope to attach the network to a chamber in the D0 experiment and investigate its performance under actual beam conditions. A muon trigger system might eventually be developed using these techniques[9]. The current D0 muon trigger system obtains an effective position resolution of about 5cm. Even if the proportional output method was used, the network could improve this by a factor of 10-20.

The distributed output technique used here to obtain the track parameters is not very efficient in the use of neurons (e.g. most of the output neurons are off in a given event.) This is a drawback for hardware implementation in that the number of available neurons on the chip is limited. Also, it obviously produces a large number of signals to manage. However, it may be quite useful to have the greater precision over the proportional net. Most importantly, though, the ability to provide secondary answers is unique to this method. This method could be used for other neural network applications, not just in hardware, where there are ambiguities and knowledge of more than one possible answer is useful.

## Acknowledgments

We give our appreciation to Gustavo Cancelo and Sten Hansen of Fermilab for their assistance. We thank Mark Holler and Finn Martin of Intel for their help. Thanks also to Giovanni Pauletta of the University of Udine. This research was supported by Fermi National Accelerator Lab.

## References

1. B. Denby *Tutorial on Neural Network Applications in High Energy Physics: A 1992 Perspective*, to be published in the proceedings of the Second International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, La Londe les Maures, France in January 1992. Also, see Denby these proceedings.
2. J. Fent, C. Kiesling, P. Ribarics, *A Level 2 Calorimeter Trigger Using Neural Networks: Status Report*, Experiment H1 internal report H1-MPI-150, 23.4.1991. Also, see Ribarics these proceedings.
3. C. S. Lindsey, B. Denby, H. Haggerty, and K. Johns, *Nucl. Inst. & Meth.*, **A317** (1992)

346-356.

4. C. Brown et al., *Nucl. Inst. & Meth.* A270 (1989) 331.
5. D. Green et al., *Nucl. Inst. & Meth.* A256 (1987) 305.
6. Intel Corp., 2250 Mission College Boulevard, MS SC9-40, Santa Clara, Ca. 95052-8125
7. M. Holler et al.,*Proc. Int. Joint Conf. on Neural Networks*, Washington, D.C., (1989), vol. II, IEEE Catalog 89CH2756, p. 191.
8. Data booklet for Intel 80170NX *Electrically Trainable Analog Neural Network*, Intel Corp., June. 1991.
9. M. Fortner, *Analog Neural Networks in an Upgraded Muon Trigger for the D0 Detector*, presented at the Second International Workshop on Software Engineering, Artificial Intelligence, and Expert Systems for High Energy and Nuclear Physics, La Londe les Maures, France in January 1992.

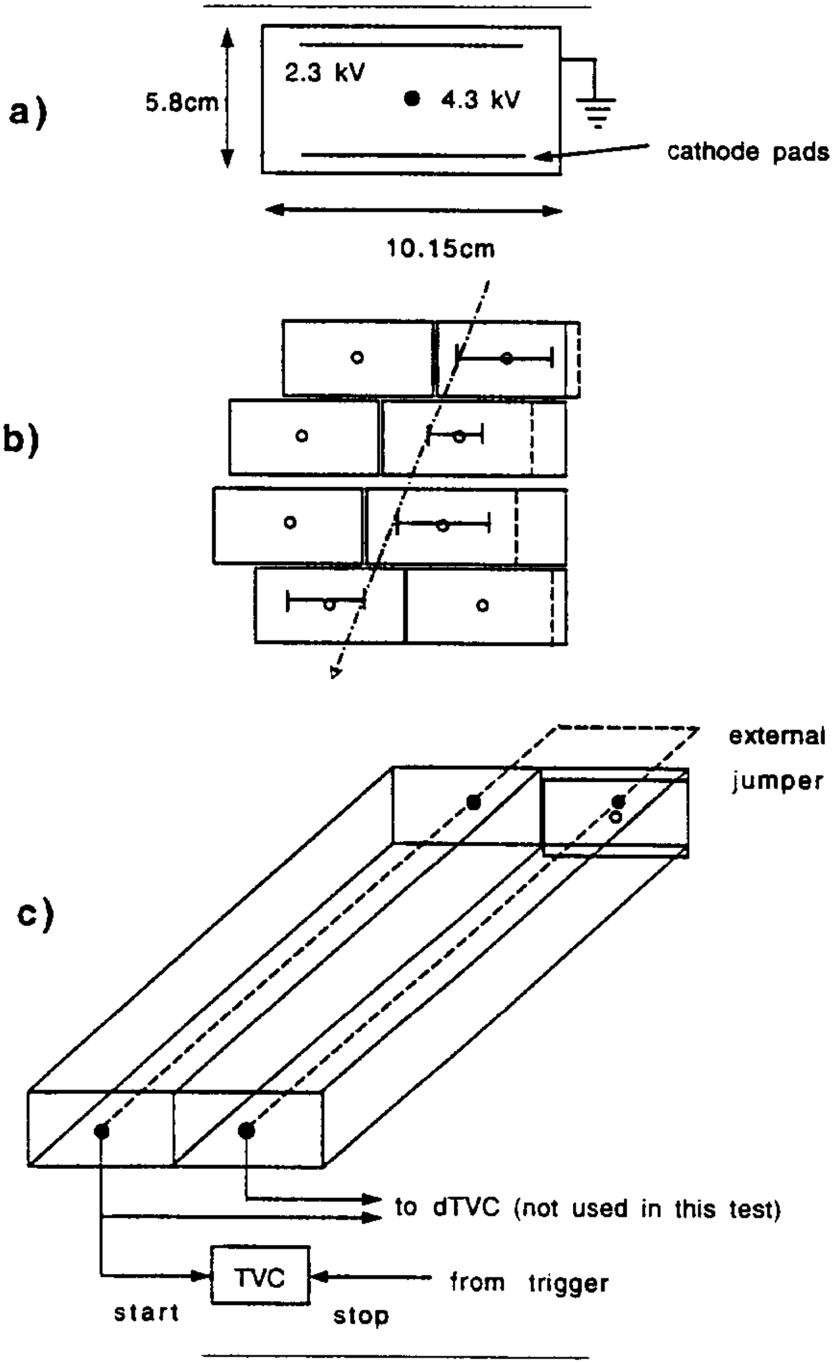

Figure 1: (a) Cross-section of a D0 muon chamber drift cell. (b) Track through a 4-layer chamber. Drift distances, with left-right ambiguities, are shown. The right edge cells are flush. Dashed lines show normal cell width. (c) Perspective view of a cell pair showing sense wire connection at one end, drift time and signal transit time electronics at other.

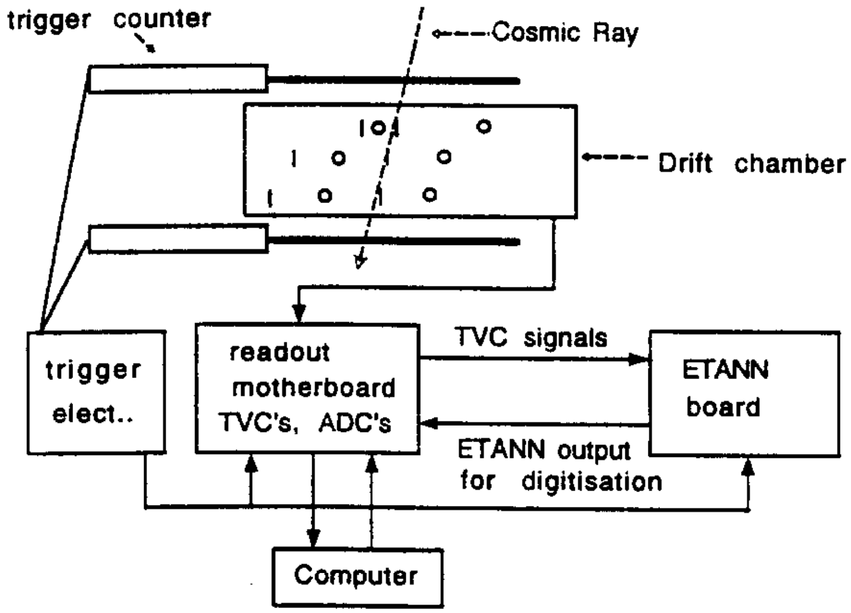

Figure 2: Test setup for tracking cosmic rays.

```
 Input  = 2 x (4 TVC Voltages) + 4 Pad Latch Voltages
Output  = 32 0.625cm bins from -0cm to +20cm
        + 32 0.05rad bins from -0.8rad to 0.8rad
```

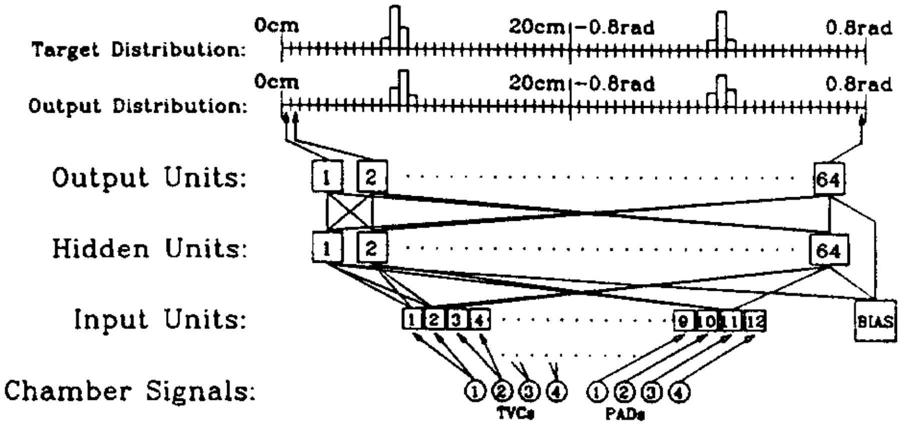

Figure 3: Two layer feedforward neural network for determining slope and intercept of tracks from drift chamber signals.

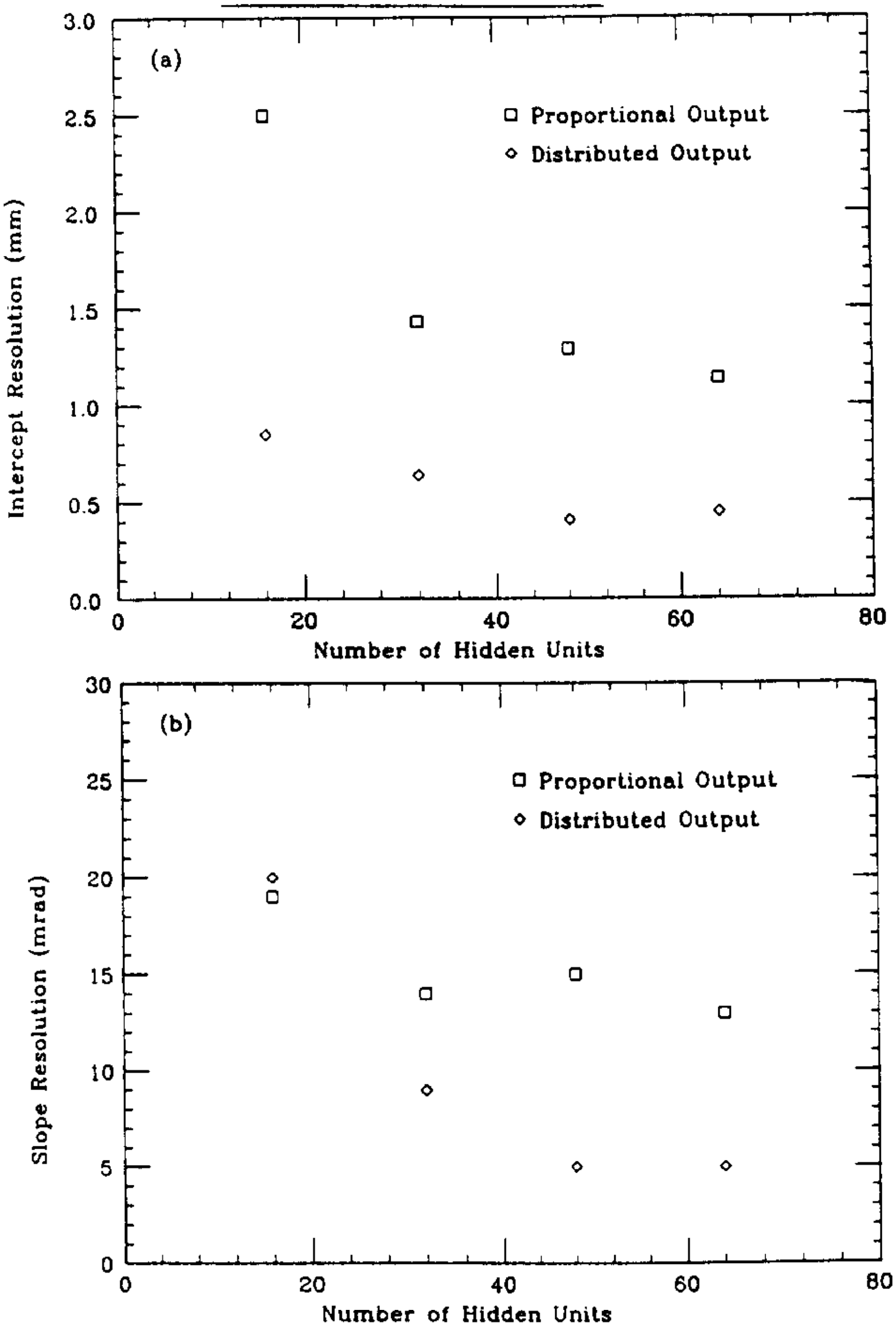

Figure 4: Resolution of (a) intercept and (b) slope using neural network simulations. Results from a net with a distributed type of output (see figure 2) are compared to results from nets where the amplitude of one output neuron is proportional to the intercept and of the amplitude of another to the slope.

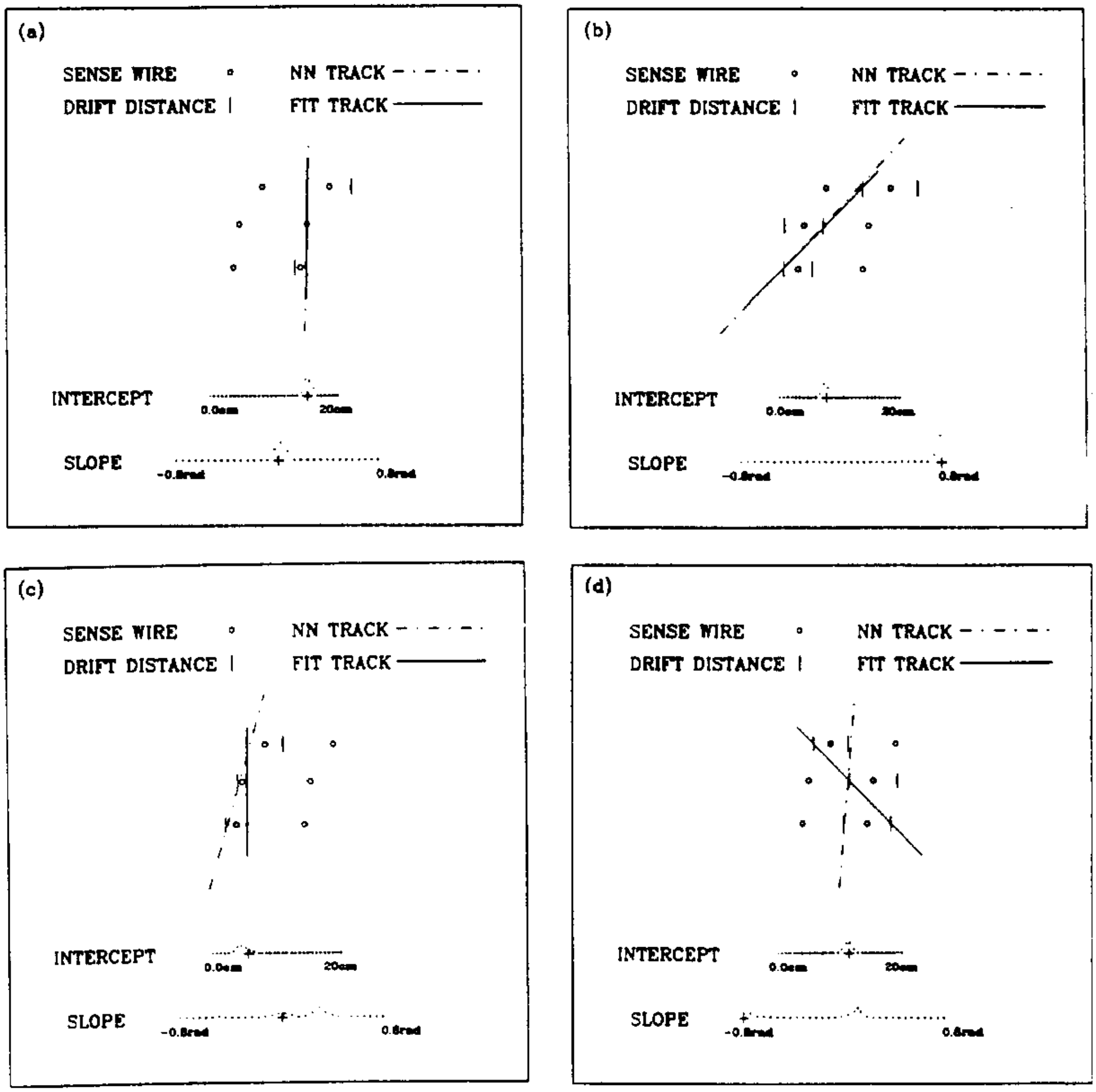

Figure 5: (a,b) Two cosmic rays through a 3-layer chamber showing the track fits and the corresponding neural network tracks. (c,d) Two tracks showing examples where neural network chooses different track than does the best fit.

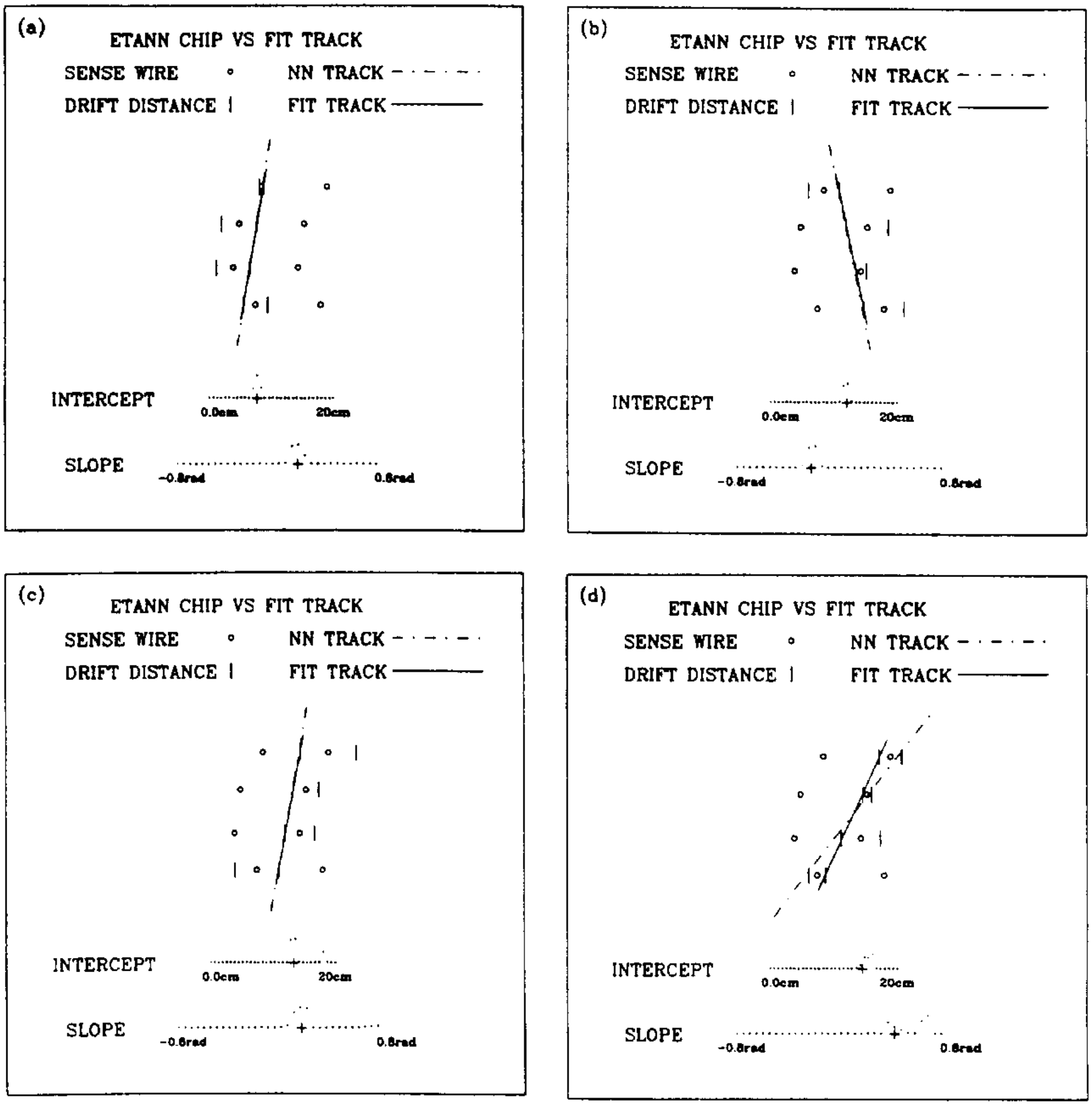

Figure 6: (a,b,c) Cosmic ray tracks through a 4-layer chamber showing the track fit and the corresponding neural network track. (d) A case where the neural network chose a different track than the best fit.